Decision Making under Uncertainty

A Multidisciplinary Perspective

11 July 2022

Contents

About

Overview

How do we make good decisions in the presence of uncertainty? This question arises in numerous contexts, including natural resources management, games, and robot planning & control. The past few decades have seen significant advances in decision-making under uncertainty. These range from new domain-independent methods in areas such as artificial intelligence, statistics, operations research, robot planning, and control theory, to novel domain-specific methods in fields such as ecology, fisheries, economics, and mathematical finance. Unfortunately, progress in one domain may often be easily overlooked by researchers from another community.

This workshop aims to provide a multidisciplinary forum for researchers from disparate fields to discuss recent advances in decision making, identify research challenges, and explore potential collaborations.

Organizers

Nan Ye, The University of Queensland
Hanna Kurniawati, Australian National University
Marcus Hoerger, The University of Queensland
Dirk Kroese, The University of Queensland
Jerzy Filar, The University of Queensland

Sponsors

The University of Queensland
Australian Research Council

Schedule

All times are in Australian Eastern Standard Time

8:30–9:00	Registration	
9:00–9:05	Opening Address: Prof. Jenny Seddon, Acting Executive Dean, FoS, UQ	
Keynote		
9:05–9:50	Alan Hájek	Ω
9:50–9:20	Coffee break	
Session A. Ubiquitous Uncertainties		
10:20–10:40	Dragan Rangelov	Perceptual Decision Making Relies on Reducing Uncertainty about Neural Sensory Representations
10:40–11:00	Antonio Rosato	Quality is in the Eye of the Beholder: Taste Projection in Markets with Observational Learning
11:00–11:20	Frankie Cho	How Uncertainty Changes Optimal Decisions for National Land Use Change
11:20–11:30	Break	
Session B. Operations Research		
11:30–11:50	Michael Forbes	An Exact Algorithm for the Pickup and Delivery Problem with Time Windows and Demand Uncertainty
11:50–12:10	Rick Jeuken	Active Set Methods for Solving Large Sample Average Approximations of Chance Constrained Optimisation Problems
12:10–12:30	Kazutoshi Yamazaki	On the CUSUM Procedure for Phase-Type Distributions: A Levy Fluctuation Theory Approach
12:30–13:30	Lunch	
Keynote		
13:30–14:15	Iadine Chadès	Developing ML and AI Decision Tools for Conservation
14:15–14:45	Coffe break	
Session C. Planning		
14:45–15:05	Marcus Hoerger	Adaptive Discretization using Voronoi Trees for Continuous-Action POMDPs
15:05–15:25	Luz Pascal	A Universal 2-state n-action Adaptive Management Solver
15:25–15:45	Nicholas Collins	Locally-Connected Interrelated Network: A Forward Propagation Primitive
15:45–15:55	Break	
Session D. Reinforcement Learning		
15:55–16:15	Jun Ju	Model-based Offline Reinforcement Learning for Sustainable Fishery Management
16:15–16:35	Vektor Dewanto	Discounting-Free Reinforcement Learning from Transient States
16:35–16:55	Konstantin Avrachenkov	Full gradient DQN Reinforcement Learning: A Provably Convergent Scheme
16.55	Closing Remarks	

List of Keynotes

Alan Hájek, Professor, Australian National University

Title: Ω

Abstract: Probability theory is the dominant approach to modeling uncertainty. We begin with a set of possibilities or outcomes, usually designated 'Ω'. We then assign probabilities — real numbers between 0 and 1 inclusive — to subsets of Ω. Nearly all of the action in the mathematics and philosophy of probability for over three and a half centuries has concerned the probabilities: their axiomatization, their associated theorems, and their interpretation. My most recent project regarding probability is to put Ω in the spotlight; this is a progress report. Ω is a set of possibilities, but which possibilities? While the probability calculus constrains our numerical assignments, and its interpretation guides us further regarding them, we are entirely left to our own devices regarding Ω. What makes one Ω better than another? Its members are typically not exhaustive — but which possibilities should be excluded? Its members are typically not maximally fine-grained — but how refined should they be? I will discuss both philosophical and practical problems with the construction of a good Ω. Along the way, I will question the notion of a 'catch-all' that is supposed to cover all possibilities that have not been explicitly identified. I will end with an omega dilemma: roughly, either orthodox probability theory breaks down, and must be rethought; or we need an account of how probabilities should be revised when a proposition that is not in Ω is learned.

Bio: Alan Hájek's research interests include the philosophical foundations of probability and decision theory, epistemology, the philosophy of science, metaphysics, and the philosophy of religion. His paper "What Conditional Probability Could Not Be" won the 2004 American Philosophical Association Article Prize for "the best article published in the previous two years" by a "younger scholar". The Philosopher's Annual selected his "Waging War on Pascal's Wager" as one of the ten best articles in philosophy in 2003.

He is a Fellow of the Australian Academy of the Humanities. He was the keynote speaker at the 2007 Chinese Analytic Philosophy Association conference, Wuhan. He was the President of the Australasian Association of Philosophy in 2009-2010. He received the 2012 Award for Excellence in Supervision, ANU College of Arts and Social Sciences. In 2013 he won the ANU-wide Vice-Chancellor's Award for Excellence in Supervision.

Iadine Chadès, Principal Research Scientist, CSIRO

Title: Developing ML and AI Decision Tools for Conservation

Abstract: I will give an overview of the ML and AI research we have been conducting to help make better decisions in the field of conservation (adaptive management) over the last 10 years. In particular, we have developed algorithms to solve and increase interpretability of Markov decision models and stochastic dynamic programming. I will be highlighting that the current bottleneck is not necessary our ability to solve complex decision problems, rather it is our ability to make solutions easy to interpret and more likely to be trusted.

Bio: Iadine's research is at the forefront of linking domain sciences such as ecology, epidemiology, synthetic biology with quantitative tools from the field of artificial intelligence (AI). She develops AI methods to provide guidance on how to make smart decisions under imperfect knowledge and resource constraints.

With CSIRO, Iadine is currently an activity leader with the MLAI Future Science Platform (2019-) where she focuses on developing ML/AI for decision-making. She is a Chief Investigator with NHMRC Centre for Research Excellence SPECTRUM (Supporting Participatory Evidence generation to Control Transmissible diseases in our Region Using Modelling). From 2012 to 2021, she was the team leader of the Conservation Decisions team (CSIRO, Land and Water) a multisdisciplinary group with expertise in ecology, systematic conservation planning, priority threat management, artificial intelligence, and decision theory.

List of Abstracts – Talks

Perceptual Decision Making Relies on Reducing Uncertainty about Neural Sensory Representations

Dragan Rangelov, The University of Queensland

Fast and accurate decisions are necessary for many adaptive behaviours. Evidence accumulation models postulate that decisions unfold gradually as evidence for different choices accumulates into an abstract decision variable. Here we modelled the content of the decision variable by manipulating uncertainty about the properties of sensory input in a speeded motion discrimination task. A group of human observers (N=36) performed the task while their brain activity was recorded using electroencephalography. Behavioural data were modelled using a drift-diffusion model, which is agnostic about the content of the decision variable and a more recent Bayesian attractor model which postulates that the decision variable accumulates uncertainty about sensory representations. The Bayesian attractor model predicted the observed data better than the drift-diffusion model, supporting the notion that the decision variable represents sensory uncertainty in simple perceptual tasks. Using multivariate analyses of observers' brain activity, neural uncertainty was estimated as time-resolved motion tuning to the presented motion stimuli. Using this estimated neural uncertainty instead of simulated uncertainty, a restricted Bayesian attractor model was fitted to the behavioural data. This model predicted the observed data as well as the unrestricted version, suggesting that perceptual decision making relies on reduction of uncertainty about neural sensory representations.

Quality is in the Eye of the Beholder: Taste Projection in Markets with Observational Learning

Antonio Rosato , The University of Queensland

We study how misperceptions of others' tastes influence beliefs, demand, and prices in a market with observational learning. Consumers infer the commonly-valued quality of a good based on the quantity demanded and price paid by other consumers. When consumers exaggerate the degree to which others' tastes resemble their own, such "taste projection" leads to erroneous and disparate quality perceptions across consumers (i.e., "quality is in the eye of the beholder"). In particular, a consumer's biased estimate of the good's quality is negatively related to her own taste. Moreover, consumers' quality estimates are increasing in the observed price, even when the price would have no influence on the beliefs of rational consumers. These biased beliefs result in perceived valuations that exhibit too little dispersion relative to rational learning and a demand function that is excessively price sensitive. We then analyze how a sophisticated monopolist optimally sets prices when facing short-lived taste-projecting consumers. Projection leads to a declining price path: the seller uses an excessively high price early on to inflate future buyers' perceptions (e.g., creating "hype"), and then lowers the price to induce a larger-than-rational share to buy. When consumers can instead time their purchase, projection causes late buyers to under-appreciate selection effects, thereby exposing them to systematic disappointment. A final application examines how projection of risk preferences distorts portfolio choice when learning from asset prices.

How Uncertainty Changes Optimal Decisions for National Land Use Change

Frankie Cho, University of Exeter & University of Queensland (QUEX Institute)

Decision making for long term land use change is inherently confounded by uncertainty over future conditions. To tame that complexity, integrated environment-economy models are frequently used to identify welfare-optimising patterns of land use change. The recommendations such models provide, however, depends on how it accommodates uncertainty. In this paper we contrast three possible strategies that handle uncertainty with increasing sophistication; (1) optimise outcome under a central future pathway (2) optimise the expected outcome across uncertain futures (3) minimise down-side risk across uncertain futures by optimising the Conditional Value-at-Risk (CVaR), a risk measure commonly used for quantifying risks in financial portfolios. We pursue that comparison in the context of a search for optimal forest planting strategies to meet the UK government's 'Net Zero' pledges. We find that how uncertainty is treated in such an exercise is critical to the recommendations it delivers, not only with regards to where to plant trees but also with regards to which species to plant. Explicit consideration of uncertainty, as per (2) and (3), delivers planting strategies that still provides a superior £8.4 to £9.4 billion natural capital benefits respectively in the worst-case scenario, compared to (1) which provides only £3.3 billion natural capital benefits in the worst-case. In the context of public decision making, adopting land use change strategies that minimise downside risk, as per (3), leads to a diversified planting strategy that hedges against scenarios with climate that do not support optimal tree growth.

An Exact Algorithm for the Pickup and Delivery Problem with Time Windows and Demand Uncertainty

Michael Forbes , The University of Queensland

We present the first exact methods for solving two stochastic variants of the Pickup and Delivery Problem with Time Windows (PDPTW). We use extended fragments, an adaptation of the method of fragments previously introduced by Alyasiry, Forbes, and Bulmer (2019). Computational results show our method is eective for the PDPTW with demand uncertainty and that it is comparable and in certain cases superior to current state-of-the-art solution techniques for the deterministic PDPTW.

Active Set Methods for Solving Large Sample Average Approximations of Chance Constrained Optimisation Problems

Rick Jeuken , The University of Queensland

This talk is about an article submitted to the Journal of Optimisation Theory and Application. We describe a novel approach to chance-constrained programming based on the sample average approximation (SAA) method. Recent work focuses on heuristic approximations to the SAA problem and we introduce a novel approach which improves on some existing methods. Our Active Set method allows one to solve SAAs of chance-constrained programs with very large numbers of scenarios quickly. We demonstrate that increasing the number of scenarios is more important than improving accuracy with small numbers of scenarios. We use an example of the portfolio selection problem to demonstrate the relative performance of previous and new methods. Extending the Active Set method to an integer-programming model further highlights its applicability and further improves over previous approaches.

On the CUSUM Procedure for Phase-Type Distributions: A Levy Fluctuation Theory Approach

Kazutoshi Yamazaki , The University of Queensland

We introduce a new method analyzing the cumulative sum (CUSUM) procedure in sequential change-point detection. When observations are phase-type distributed and the post-change distribution is given by exponential tilting of its pre-change distribution, the first passage analysis of the CUSUM statistic is reduced to that of a certain Markov additive process. By using the theory of the so-called scale matrix and further developing it, we derive exact expressions of the average run length, average detection delay, and false alarm probability under the CUSUM procedure. The proposed method is robust and applicable in a general setting with non-i.i.d. observations. Joint work with J. Ivanovs (Aarhus University).

Adaptive Discretization using Voronoi Trees for Continuous-Action POMDPs

Marcus Hoerger , The University of Queensland

Decision making under partial observability is an essential capability of autonomous robots to complete a given task robustly. The Partially Observable Markov Decision Processes (POMDP) is a principled framework that enable autonomous robots to make good decisions in the presence of various uncertainties. While POMDPs are notoriously difficult to solve exactly for all but the simplest planning problems, the past two decades have brought tremendous advances in developing approximate solvers, making them viable tools for realistic planning problems under partial observability. Despite this progress, continuous action spaces remain a fundamental challenge in POMDP planning. In this talk I will present a new online POMDP solver, called Adaptive Discretization using Voronoi Trees (ADVT), designed to handle continuous action spaces more effectively compared to state-of-the-art methods. At its core, ADVT combines Monte Carlo Tree Search (MCTS) with a novel adaptive action-discretization method called Voronoi Tree. Our proposed method aims to better utilize local information in the action space during online planning, which helps ADVT in scaling to higher-dimensional action spaces more efficiently, compared to existing methods.

A Universal 2-state n-action Adaptive Management Solver

Luz Pascal , Queensland University of Technology

In poor data and urgent decision-making applications, managers need to make decisions with incomplete knowledge of the system dynamics. In biodiversity conservation, adaptive management (AM) is the recommended practice for decision-making under uncertainty. hidden model MDPs (hmMDPs), a simplified Mixed Observable Markov Decision Processes, provide optimal solutions to AM problems when the dynamics of the system are unknown. hmMDPs augment the MDP state space with an unobservable state variable representing a predefined set of possible dynamics of the system. A drawback in formalising an AM problem as a hmMDP is the assumption that the real system dynamics are included in the predefined set. Current methods rely on expert elicitation, which is a challenging and time-consuming process that is prone to biases. We propose an new approach to build a hmMDP with a universal set of predefined models that is capable of solving any 2-state n-action AM problem. Our approach uses properties of the transition matrices to build the model set. Our approach is fast, robust and independent of expert input. It relies on analytical formulations to derive the minimum set of models to include into an hmMDP to solve any AM problems with 2 states and n actions. We assess our universal AM algorithm on two species conservation case studies from Australia and randomly generated problems.

Locally-Connected Interrelated Network: A Forward Propagation Primitive

Nicholas Collins , The University of Queensland

End-to-end learning for planning is a promising approach for finding good robot strategies in situations where the state transition, observation, and reward functions are initially unknown. Many neural network architectures for this approach have shown positive results. Across these networks, seemingly small components have been used repeatedly in different architectures, which means improving the efficiency of these components has great potential to improve the overall performance of the network. This paper aims to improve one such component: The forward propagation module. In particular, we propose Locally-Connected Interrelated Network (LCI-Net) —a novel type of locally connected layer with unshared but interrelated weights— to improve the efficiency of learning stochastic transition models for planning and propagating information via the learned transition models. LCI-Net is a small differentiable neural network module that can be plugged into various existing architectures. For evaluation purposes, we apply LCI-Net to VIN and QMDP-Net. VIN is an end-to-end neural network for solving Markov Decision Processes (MDPs) whose transition and reward functions are initially unknown, while QMDP-Net is its counterpart for the Partially Observable Markov Decision Process (POMDP) whose transition, observation, and reward functions are initially unknown. Simulation tests on benchmark problems involving 2D and 3D navigation and grasping indicate promising results: Changing only the forward propagation module alone with LCI-Net improves

VIN's and QMDP-Net generalization capability by more than 3× and 10×, respectively.

Model-based Offline Reinforcement Learning for Sustainable Fishery Management

Jun Ju , The University of Queensland

Reinforcement learning provides a simple and general computational approach for sequential decision making. Recent advances of reinforcement learning focus on fully observable environments. However, real-world problems are often partially observable, and it remains a challenge to develop sample-efficient and robust algorithms for such problems. In this talk, I will first show how to address these challenges in an important partially observable domain, sustainable fishery management. Specifically, we propose MOOR, a robust and interpretable model-based offline reinforcement learning algorithm. MOOR achieves sample-efficiency by first learning fishery POMDP models, then learning a strategy from past fishery data without directly interacting with the environment. We simulated noisy data and incomplete data sequences for the experiments. For each dataset, we try to learn POMDPs using both well-specified and misspecified models. The simulation study show that MOOR can learn a fairly accurate model and achieve similar level of policy values as using the ground truth model. In addition, I will introduce an interpretable policy, which we call the threshold policy. Our initial experiments on deterministic data also show such policies are robust against model learning errors. Finally, I will provide a theoretical justification of the effectiveness of threshold policies.

Discounting-Free Reinforcement Learning from Transient States

Vektor Dewanto , The University of Queensland

Most reinforcement learning methods (RL) optimize a discounted-reward objective, even for environments without any inherent notion of discounting. An alternative approach is to optimize an average-reward objective, which is discounting-free such that several complications due to artificial discounting can be avoided. However, the average-reward objective focuses only on long-run rewards. It therefore neglects rewards earned at the outset, particularly those earned in the transient states. I would like to talk about our new discounting-free policy gradient RL technique that is able to yield approximately optimal policies with respect to the (most selective) Blackwell's criterion. This property holds not only in recurrent but also transient states. We derive a gradient expression of the bias that enables approximation by sampling. We also propose a preconditioning matrix for the gradient along with its sampling-enabler expression. We further devise an algorithm that utilizes the proposed estimators. Experimental results provide insights into the fundamental mechanisms of our proposal.

Full gradient DQN Reinforcement Learning: A Provably Convergent Scheme

Konstantin Avrachenkov , INRIA Sophia Antipolis

We analyze the DQN reinforcement learning algorithm developed by DeepMind as a stochastic approximation scheme using the o.d.e. (for 'ordinary differential equation') approach and point out certain theoretical issues. We then propose a modified scheme called Full Gradient DQN (FG-DQN, for short) that has a sound theoretical basis and compare it with the original scheme on benchmark problems. We observe a better performance for FG-DQN. This is a joint work with V.S. Borkar, H.P. Dolhare and K. Patil.

Notes

Notes

Notes